Pillole di false IA 631/ La bolla finanziaria degli LLM

(631)—Riconoscere che l'attore principale dello sviluppo delle false IA è la finanza è solo il primo passo per porsi le domande giuste…

Pillole di false IA 631/ La bolla finanziaria degli LLM



Figure 1:

(631)—Riconoscere che l'attore principale dello sviluppo delle false IA è la finanza è solo il primo passo per porsi le domande giuste sul futuro che ci aspetta.

5 agosto 2025—In queste "pillole" Cassandra ha cercato di dividere in due le sue esternazioni agostane, trattando prima il piano tecnologico, ed in quella di oggi il piano finanziario.

Oggi tocca appunto alla finanza.

Se c'è una cosa che la fa più arrabbiare di come vengono usati gli LLM è come vengono manovrate le montagne di soldi che, con la scusa degli LLM, vengono spesi ed investiti da ormai quasi tre anni, e che potrebbero servire per scopi più banali ed importanti, e rendere il mondo un posto un po' migliore.

Cassandra si è accorta di non essere in questo da sola; un esempio di persona altrettanto arrabbiata sul tema LLM e finanza, è Edward Zitron, che in questo articolo scrive paginate di

dati e dettagli su ciò che pensa della "Bolla" finanziaria degli LLM. Cassandra ha cominciato a farlo prima di lui, quindi lo sente come un fratello minore.

Pur senza addentrarsi in dati finanziari, il ragionamento (chiamiamolo così) di Cassandra viaggia su un binario assolutamente equivalente e parallelo.

Tutto nasce dal ricordo chiarissimo di quella che è stata la bolla delle dot.com, sviluppatasi e poi scoppiata dal 1997 al 2000. Chi non l'ha vissuta, e non ne è stato avvolto, subendone suo malgrado gli effetti, non può percepire a pelle le assonanze con l'attuale e molto più grande bolla degli LLM.

Allora, Cassandra si ricorda perfettamente, successero cose che a riguardarle oggi sembrano incredibili. Ad esempio Tiscali, in un certo periodo, valeva in borsa più della Fiat, allora ancora in vita. La stampa di settore approvava, osannava e prediceva un futuro radioso e progressivo,

E la narrazione delle dot.com sembrava credibile alla maggioranza delle persone, inclusa la non ancora svezzata Cassandra.

Ecco, vediamo di mettere giù alcuni punti fermi a cui agganciare similitudini e differenze tra il 2000 ed il 2025. Facciamoci domande semplici per dare risposte semplici.

Esiste una bolla delle IA?

Evidentemente si, ma non è corretto chiamarla così; la bolla è quasi completamente causata dagli LLM, ed è quindi più preciso chiamarla "*Bolla degli LLM*".

Quanto è grande la Bolla degli LLM rispetto a quella delle dot.com?

Difficile calcolarlo. Possiamo ricordare che non esistevano all'epoca, dot.com che capitalizzassero più di un trilione di dollari, mentre oggi ce ne sono otto.

La più capitalizzata, NVIDIA, vale da sola più di 4 trilioni, ed è l'unica azienda che guadagna soldi veri dal fenomeno degli LLM. Vende chip per l'elaborazione grafica dei videogiochi, modesto mestiere, fatto in passato in condizioni di quasi monopolio, ma navigando allora su quotazioni di frazioni di dollaro.

Quando la bolla degli LLM ha cominciato a gonfiarsi, ci si è resi conto che i calcoli degli LLM potevano essere fatti molto più velocemente utilizzando questi chip grafici. I chip sono diventati introvabili, le quotazioni dell'azienda sono aumentate di 200 volte, rimanendo sostanzialmente la stessa azienda; pur producendo gli stessi chip in volumi simili.

Vi sembra sano o strano?

Le 5 aziende successive in capitalizzazione, Apple, Amazon, Google, Meta e Microsoft sono decisamente delle dot.com. Due di esse (Google e Meta) regalano i loro prodotti principali e guadagnano vendendo dati, mentre le altre 3, pur non disdegnando questo mercimonio, producono principalmente software.

I veri produttori di LLM, invece, non sono quotati in borsa, ma operano ricevendo finanziamenti a fondo perduto e partecipazioni azionarie da altre aziende. Parliamo di OpenAI, Antrophic, etc.

Possiamo notare che solo NVIDIA vende un prodotto reale, ed è infatti, fornitore di OpenAI, Anthropic, Google, Meta, tutti insomma. Data la produzione limitata di chip, i suoi prezzi sono altissimi ed il fatturato di NVIDIA è incommensurabilmente più basso di quello che la sua capitalizzazione, se fosse organica, richiederebbe.

Cosa succederà appena i cinesi od altri produrranno in massa chip equivalenti e NVIDIA tornerà a valere quello che vale? Una catastrofe finanziaria?

Veniamo ai Signori dell'AI.

OpenAI, Anthropic e gli altri fornitori di LLM sono tutti consumatori netti di capitali, con fatturati nulli o molte volte inferiori ai soli costi operativi che sostengono. Ambedue hanno perso, nell'ultimo anno fiscale 2024, oltre 4 miliardi, e malgrado questo continuano ad attrarre investimenti di decine di miliardi di dollari, promettendo agli investitori guadagni non prima della fine del decennio.

Ricordiamo che gli LLM sono prodotti venduti per quello che non sono e per quello che non possono fare.

Sono oggetto di campagne di marketing e di informazione ingannevole di intensità e sofisticazione tali da non avere precedenti nella storia.

Si continua a sostenere che gli LLM produrranno redditi e risparmi miliardari in moltissimi ambiti, ma per ora non ci sono storie di successo, anche piccole, che avvalorino queste promesse. Sam Altman, CEO di OpenAI, ha dichiarato che gli servono 40 miliardi di dollari nei prossimi tre anni per mantenere le promesse fatte. E pare che li abbia anche trovati.

Non ci sono casi di successo aziendali dell'IA, né se ne prevedono a breve.

Tuttavia tutte le aziende che hanno investito non nei giganti dell'IA ma direttamente negli LLM stanno infilandoli in qualsiasi prodotto vendano.

Fanno questo per motivi pubblicitari? In parte si, ma il motivo principale è quello di tentare di generare ritorni finanziari per gli investimenti fatti, a spese ovviamente del loro parco utenti.

Lo fanno facendo pagare un modesto abbonamento (es. OpenAI) oppure inventandosi un nuovo prodotto od espandendone uno esistente, aumentandolo di prezzo (Microsoft).

Ma qualche centinaio di dollari per utente sono una goccia nel mare di investimenti fatti durante l'espansione della Bolla degli LLM, che sono investimenti pronti a trasformarsi in perdite. Includono anche gli investimenti indiretti in centri di calcolo enormi, e centrali elettriche che oggi non esistono e che dovranno alimentarli, cosa che per ora avverrà sottraendo potenza agli impieghi ordinari.

Pare proprio che l'industria degli LLM sia controllata da chi vuole operare nella finanza ma non ha nessun interesse a farlo in maniera sana, e per costruire un futuro migliore a misura d'uomo.

Ma alla fine chi usa l'IA, veramente? Sono singole persone che usano malamente gli LLM per aumentare la loro produttività a scapito della qualità e dell'affidabilità

Evitiamo per un attimo i paralleli con la storia. Evitiamo pure la discussione su cosa sia l'intelligenza, e come la si possa eventualmente far simulare ad un computer.

Parliamo per un attimo solo di LLM per riassumere la questione. Gli LLM non vengono testati con metodi scientifici, vengono pubblicizzati; basta vedere gli annunci che ne vengono continuamente fatti.

Un LLM non apprende; il modello fondativo viene allenato fornendo tutto lo scritto esistente al mondo (parliamo degli LLM generalisti come Claude o ChatGPT). L'informazione che l'LLM elabora viene distrutta e produce una configurazione di parametri interna alla struttura dell'LLM stesso (il modello fondativo).

Il blob di n GigaByte che scaricate per installare Llama sul vostro pc, ad esempio, è esattamente questo. Poi gli LLM commerciali vengono "allineati" di solito inserendo circuiti censori in input ed in output, che sono o semplici filtri od altri LLM che agiscono in maniera adversarial.

Dentro il modello fondativo non c'è più il testo usato per allenarlo, ci sono solo coefficienti statistici di probabilità della sequenza dei token.

Gli LLM, tutti gli LLM funzionano così.

I discorsi che ci vengono costruiti sopra da chi li descrive, li vende o li usa sono dei castelli in aria per gonfiare o tenere gonfia la Bolla degli LLM.

La produzione che si ottiene da un LLM dopo averlo ulteriormente nutrito di un prompt e di una "domanda" non è una "risposta", ma una sequenza di parole valutata in termini di probabilità.

Non c'è ragionamento, non c'è informazione. Gli LLM sono solo elaboratori meccanici di linguaggio naturale. Questo è il loro mestiere e lo sanno fare benissimo. Servono ad elaborare il linguaggio, tradurre, cambiare lo stile, tentare di riassumere.

Oggi invece vengono usati per affabulare ed affascinare, o peggio come oracoli da interrogare, e come mezzo per creare la bolla finanziaria di cui vedremo gli effetti tra poco. Si invogliano le persone ad utilizzarli per aumentare la loro produttività e risparmiare tempo.

Tutto questo viene ottenuto diminuendo la qualità e la novità di ciò che viene prodotto. Talvolta inserendo errori, enormi o molto sottili, in ciò che viene prodotto. Non per nulla gli LLM collassano se vengono allenati con testi prodotti da altri LLM.

Ma forse siete di quelli che credono nelle "*Proprietà emergenti degli LLM*" come garanzia per il futuro radioso dell'IA? A Cassandra viene quasi da piangere!

Stiamo facendo del male, molto male, non tanto a noi stessi, ma al complesso delle conoscenze (Cassandra lo chiama "Infosfera", se volete chiamiamolo cultura o quello che preferite) inquinandolo in maniera probabilmente irrimediabile. Se fate un giro su Arxiv, fate il revisore di articoli per una rivista oppure il peer reviewer per una pubblicazione accademica, ve ne accorgete immediatamente.

Cosa sta succedendo ultimamente? Le aziende che hanno investito non vedono ritorni, e nemmeno ragioni per cui questi ritorni dovrebbero arrivare nel prossimo futuro, e cominciano ad esitare. La Bolla inizia a tremare. Come evitare di perdere gli investimenti?

Ecco che arrivano nuovi ed opportuni annunci mirabolanti. Altman ed Ives annunciano il prossimo grande prodotto per interfacciarsi con l'IA del futuro, senza dire nemmeno di cosa si tratti. La corsa all'AGI torna ad essere lo spirito guida per convincere gli investitori, come in una seduta spiritica, che un aldilà lucroso esiste.

Si tenta insomma di gonfiare una nuova bolla nella Bolla degli LLM per sostenerla.

Cosa potrebbe mai andare storto?

Anche cose oggettivamente positive potrebbero bastare a far scoppiare la Bolla degli LLM. Dove sta scritto che il prossimo vero progresso nel campo dell'IA avrà la fame smodata di calcoli che hanno gli LLM? Magari basterà uno smartphone.

Dove sta scritto che i prossimi chip per l'IA costeranno carissimi? Come per DeepSeek, ad esempio, alla Cina basterebbe un annuncio con un fondo minimo di realtà per far saltare il banco e scoppiare la Bolla degli LLM.

Un atto di guerra asimmetrica, come ce ne sono già stati tanti da parte di tutti nella storia della finanza, che stavolta potrebbe colpire il ventre molle della Bolla degli LLM.

Cassandra sarebbe curiosa di sapere cosa ne pensate.

Buone vacanze.

Scrivere a Cassandra—Mastodon
Videorubrica "Quattro chiacchiere con Cassandra"
Lo Slog (Static Blog) di Cassandra
L'archivio di Cassandra: scuola, formazione e pensiero

Licenza d'utilizzo*: i contenuti di questo articolo, dove non diversamente indicato, sono sotto licenza Creative Commons Attribuzione—Condividi allo stesso modo 4.0 Internazionale (CC BY-SA 4.0),* tutte le informazioni di utilizzo del materiale sono disponibili a questo link*.*

By Marco A. L. Calamari on August 7, 2025.

Canonical link

Exported from Medium on August 27, 2025.